Journal of Education for Teaching, Vol. 29, No. 1, 2003

Carfax Publishing
Taylor & Francis Group

Getting in Step: accountability, accreditation and the standardization of teacher education in the United States

ROBERT V. BULLOUGH, JR., D. CECIL CLARK & ROBERT S. PATTERSON
CITES (Center for the Improvement of Teacher Education and Schooling), Brigham Young University, Provo, Utah, USA

ABSTRACT *The background to the 'standards movement' in education in the United States is explained. It is argued that it has had a negative impact on teacher education and teacher education accreditation. An illustrative example is provided from a recent accreditation review of the application of an overly narrow understanding of standards and of legitimate evidence, and a possible alternative and more responsive and responsible view of assessment is suggested. The value of standards as guidelines for improving practice in teacher education is not questioned. Rather we aim to cast light upon what we see as their troubling reduction to templates under the often unrecognized influence of market assumptions about teaching and learning.*

INTRODUCTION: TEACHER EDUCATION ACCREDITATION: NCATE

Thirteen hundred higher education institutions prepare teachers to teach in the United States. Of these, roughly half are accredited by NCATE, the National Council for the Accreditation of Teacher Education (www.ncate.org), which was founded in 1954. Recognized by the US Department of Education, NCATE is composed of representatives from more than 30 national professional associations. Thus, unlike in England and Wales where teacher education is subjected to periodic external review by government inspectors from the Office for Standards in Education (OfSTED) (see Gilroy, 1999; Poppleton, 1999; Universities Council for the Education of Teachers, 1999), peer review remains in place in the USA. Among the non-accredited teacher education institutions in the USA are some of the most distinguished teacher education institutions in the nation as well as large numbers of small, liberal arts, colleges that often employ just a few teacher educators who alone carry the burden of certification.

The aim of accreditation, as a form of accountability, is to provide 'assurance to the public

ISSN 0260-7476 print; ISSN 1360-0540 online/03/010035-17

DOI: 10.1080/0260747022000057945

that the graduates of accredited institutions have acquired the knowledge, skills, and dispositions necessary to help all students learn' (NCATE, 2002, p. 1). Quality teachers are to be the result. Institutions wishing to be accredited by NCATE gather a range of materials and write a self-study in preparation for a visit by a committee of from three to eight members (a Board of Examiners) that spends a few days on campus reviewing the 'professional education unit [including] all initial teacher preparation and advanced programs offered by the institutions for the purpose of preparing teachers and other professional education personnel to work in pre-kindergarten through twelfth grade settings' (NCATE, 2002, p. 5). After its campus visit, the Board of Examiners writes a unit review which is forwarded to NCATE's Unit Accreditation Board which makes a final recommendation. A unit wishing to maintain accreditation will be reviewed every 5 years, but yearly reports are required.

While historically accreditation has been voluntary, the past few years there has been a persistent movement in the various states to require it. Soon, some form of national accreditation, in addition to already existing individual state accreditation programs, in all likelihood will be required. To this end, there has been a strong movement toward coordinating state and national accreditation and toward articulation of accreditation standards. Given the power of these trends, it is imperative the process be critically examined.

Background

Since 1983, and the publication of *A Nation at Risk* (National Commission on Excellence in Education) which declared publicly funded education in the US a failure, educators have increasingly confronted a frustrated and often angry public. Stirred by politicians who are quick to blame the schools for every imaginable social ill from high rates of teenage pregnancy to declining civility, Americans are demanding new levels of teacher accountability and supporting development and implementation of set and measurable standards of performance that enable comparison across programs, schools, and states. In some respects the situation is like that in England described by Poppleton (1999): 'When governments become dissatisfied with the output of their schools, they seek to increase direct curriculum control, monitor teacher quality and enhance the accountability of schools' (p. 235). Proof of the value of the investment in schooling in the form of clear social and economic dividends is required.

Although delayed, similar pressures now firmly bear down upon teacher education. In both publicly funded schooling and in teacher education, accountability is the watchword: 'To call for accountability is to assert a political right—to demand that a particular individual or institution assume some responsibility and demonstrate it in a certain form' (Smith & Fey, 2000, p. 335). There is general agreement about educational reform: the key is 'closer and closer alignment of standards for teaching and learning, assessments of students and teachers, and new models of teacher education, licensing, and certification' (Cochran-Smith, 2001, p. 538).

Encouraged by recent legislation championed by President Bush ('No Child Left Behind' Act of 2001), American children, already among the most tested in the Western world, are being tested more and scores on standardized tests are being published and school faculties rewarded and punished based upon test results (see Linn *et al*., 2002). Competition is accepted as the road to achieving heaven on Earth. Schools, like children, are being graded and ranked (see Keller, 2002). Nelson (2002) nicely captures the tenor of the times:

> The current conservative political climate is having a devastating effect on those involved in the education of our children. Much of the political language focusing on schools and education today is a language of failure rather than a language of success. Educational accountability for teachers, schools, and students appears directed toward identifying those not achieving under highly prescriptive standards as failures and prompts an even more strong handed, top–down decision-making process that tends to further exacerbate the problem. Federal and state funds are withheld, teachers and administrators reel with stress, and families and children are caught in a wave of uncertainty. What we are witnessing in today's educational climate is far from hopeful, primarily because of the ways in which legislated mandates are used to measure, prescribe, reward, and punish those educators working with our nation's youth. Just today a news report described the suicide of a school principal in Atlanta as being related to her school coming up one-tenth of a point short on a series of standardized tests measuring school achievement. Her school was identified as failing and subject to drastic measures including busing students to distant schools and dismantling the faculty. Teachers and administrators are currently under siege … (p. 3)

Under such pressures innovative approaches to evaluation, to providing evidence of accomplishment, have given way to more standardized and traditional models of testing. Developments in the state of Maryland provide an example: over a period of several years a complex system for testing children was developed for the purpose of gathering data useful for making judgments about the quality of individual school curricula that involved writing essays, laboratory reports, and engaging in mathematical reasoning. Test results helped educators improve programs. This system will now be replaced with one representing 'traditional testing practices that cover a breadth of material and are easy to score' (Hoff, 2002, p. 27). Indeed, it is likely that the ' "No Child Left Behind" Act of 2001 will have a significant impact on state testing systems throughout the country, possibly pushing many states toward the norm-referenced tests that are quick and easy to administer and score' (Hoff, 2002, p. 27). It is not only the President and other politicians who support the shift, however, but also many educators and anxious parents as well who want simple measures of individual children's learning. Tests of this kind are easy to understand and their purposes now are taken as common sense—to rank schools and to evaluate children.

Growing Disappointment

Disillusionment with schooling has many sources: distrust of teachers and of intellectuals; politicians who pander to and exploit the public's fears of cultural dissolution and who

loudly proclaim the schools to be hotbeds of immorality and centers of relativity and mediocrity; a desire to undermine publicly funded education with the aim of gaining public monies for private schooling in the form of vouchers or tuition tax credits; economic uncertainty and the failure of a diploma to assure employment; and a century-long shift of social responsibilities to the schools and away form other social institutions that inevitably produced a loss of what was a naive faith (see Tyack & Cuban, 1995). James Gallagher (2002) offers yet another reason: 'inflated expectations of consumers as to what can or should be accomplished through educational innovation and research' (p. 52). Used to and spoiled by fantastic technological breakthroughs arising from basic science, simple solutions for complex social problems are expected by Americans and when they are not forthcoming disillusionment follows and teachers and eventually teacher educators are blamed. There is irony here: 'teachers and teacher educators are being constructed as both the last great hope and the most culpable culprits in what ails American schools' (Cochran-Smith, 2001, p. 540).

Reflecting a kind of 'technocratic mindedness' (Bullough *et al*., 1984), quick, permanent, and relatively inexpensive fixes to what are in actuality intractable problems are sought. The public is not satisfied with what education and other social science researchers can provide, namely, means for achieving incremental gains, slow improvement on the margins. Facing disappointment amid growing uncertainty, rising test scores provide parents means for gaining a small measure of comfort about their children's future, educators a bit of evidence that their hard work has some sort of benefit, and politicians confirmation that they were wise to impose external controls on teachers and to limit their influence and the influence of teacher educators over policy.

A Shift in Concern

Until the late 1990s, the central aim of teacher education accreditation in the USA was to ensure that institutions had a clear focus and sufficient resources to provide a quality program and that each intending teacher had been exposed to the subject matter and pedagogical content and had a range of experiences of a specified duration in schools deemed essential to learning to teach. Inputs mattered most. This has changed somewhat. Inputs still matter, but demonstrated performance is the central concern, and to this end NCATE, claiming to speak for 'the profession', has developed a new set of standards.

> The standards measure an institution's effectiveness according to the profession's expectations for high quality teacher preparation as America enters the 21st century. The profession of teaching has developed and articulated standards for the preparation of those who enter its ranks. The profession has reached a consensus about the knowledge and skills a teacher needs to help P-12 students learn. That consensus forms the basis for the NCATE standards (NCATE, 2002, p. 7).

To be accredited a unit must provide evidence in the self-study that its programs produce desired changes in candidate knowledge, skill and disposition, which together are deemed

'learning outcomes'. What counts as evidence and how evidence is marshaled to make a case for quality have become crucial questions.

Six standards have been identified, which include 'rubrics' (criteria) that delineate the elements of each standard and describe three proficiency levels ('unacceptable', 'acceptable', and 'target') at which each element is being addressed and a 'descriptive explanation of the standard' (NCATE, 2002, p. 8). Sitting over the six standards is a unit's 'conceptual framework' which is to establish 'the shared vision for a unit's efforts in preparing educators to work effectively in P-12 schools [and which] provides direction for programs, courses, teaching, candidate performance, scholarship, service, and unit accountability' (NCATE, 2002, p. 10). Significantly, NCATE wording suggests that a unit might have multiple conceptual frameworks.

The six standards are:

1. candidate knowledge, skills and dispositions: 'Candidates preparing to work in schools as teachers … know and demonstrate the content, pedagogical, and professional knowledge, skills, and dispositions necessary to help all students learn. Assessments indicate that candidates meet professional, state, and institutional standards' (NCATE, 2002, p. 10);
2. assessment system and unit evaluation: 'The unit has an assessment system that collects and analyzes data on the applicant qualifications, candidate and graduate performance, and unit operations to evaluate and improve the unit and its programs' (NCATE, 2002, p. 10);
3. field experiences and clinical practice: 'The unit and its school partners design, implement, and evaluate field experiences and clinical practice so that teacher candidates … develop and demonstrate the knowledge, skills, and dispositions necessary to help all students learn';
4. diversity: 'That unit designs, implements, and evaluates curriculum and experiences for candidates to acquire and apply the knowledge, skills, and dispositions necessary to help all students learn …' (NCATE, 2002, p. 10);
5. faculty qualifications, performance, and development: 'Faculty are qualified and model best professional practice … The unit systematically evaluates faculty performance and facilitates professional development' (NCATE, 2002, p. 11); and
6. unit governance and resources: 'The unit has the leadership, authority, budget, personnel, facilities, and resources … for the preparation of candidates to meet professional, state, and institutional standards' (NCATE, 2002, p. 11).

As noted, each standard has a set of rubrics attached to it which 'are to be interpreted holistically; they are not to be used to make a separate judgment on each element of a standard' (NCATE, 2002, p. 11). The rubrics for Standard I, candidate knowledge, skills, and dispositions, are as follows:

- unacceptable: 'Teacher candidates have inadequate knowledge of subject matter that they plan to teach and are unable to give examples of important principles or concepts delineate in professional, state, and institutional standards';

- acceptable: 'Teacher candidates know the subject matter that they plan to teach and can explain important principles and concepts delineated in professional, state, and institutional standards';
- target: 'Teacher candidates have in-depth knowledge of the subject matter that they plan to teach as described in professional, state, and institutional standards. They demonstrate their knowledge through inquiry, critical analysis, and synthesis of the subject' (NCATE, 2002, p. 14).

Assessment and the Problem of Grace

Despite representing a separate standard (Standard II), assessment cuts across each of the standards and, as noted, is presumably linked tightly to the conceptual framework. In the glossary to its standards, an 'Assessment System' is defined as being a 'comprehensive and integrated set of evaluation measures that provides information for use in monitoring candidate performance and managing and improving unit operations and programs for the preparation of professional educators' (NCATE, 2002, p. 52). This system is central to what NCATE claims is its driving purpose, to encourage 'continuous improvement [in teacher education] based on accurate and consistent data' (NCATE, 2002, p. 1). Measurement is assumed to be the key to improvement and proof of improvement is measured performance. The word 'performance' is so weighty that it comprises no fewer than seven terms ('performance data') defined in the glossary.

Recognizing the difficulty of shifting from an input to an output accreditation model and of designing then implementing an assessment system, a transition period was established by NCATE. Institutions whose regular 5-year accreditation review fell within this period would be judged in slightly different ways from those that would be visited later, after the grace period ended. For those falling within the transition period, evidence of good faith in moving towards achieving the new standards presumably would be sufficient for accreditation. It is expected that data will follow.

MAKING SENSE OF ASSESSMENT IN THE QUEST FOR OUTCOMES

A great deal has been written on assessment. Few if any educators would argue against the importance and value of accountability; educators expect to be held accountable. However, a lively debate surrounds the question of how accountability is to be established and about the place and value of professionalism in accountability. Drawing on recent developments in business, Powell (2000) makes a case for the value of self-assessment, a key component of professionalism, and presents a compelling criticism of traditional evaluation practices and their concern with outcome measures: 'There has been little understanding of how the internal culture of the school works and how the use of a wider range of assessment tools, such as self-assessment, can facilitate improvement strategies. There has been a consistent failure to question the nature of assessment in relation to the qualities it does or should identify and influence, not only for pupils, but also for teachers' (p. 39). She suggests that the nature of change in complex organizations argues against traditional

evaluation practices that denigrate the individual teacher by centralizing policy and resources and by holding teachers responsible to standards over which they have no influence and by decreasing the range of their professional decision-making. Such impositional models are grounded in distrust, and reflect a domineering instrumental interest in control. Externally imposed, such models seek to compel rather than invite change and development.

Aims, as Dewey would remind us, must not be separated from means. When they are separated, when aims lose their human grounding, they lose their ability to inspire action and stimulate imagination. When means become aims, journeys become long and pointless and life weary. Just so with accreditation: when compliance with outcomes becomes the aim of accreditation, program quality suffers and the arena of human professional action is truncated and growth limited. As Cochran-Smith (2001) observes, notwithstanding NCATE, 'there is not consensus in the US about how and where teachers should be educated, what they should learn (or not learn), and what theories of teaching and learning should guide their learning' (p. 538). Outcomes are expressions of political understanding, and as such must be publicly contested. The irony is that the shift toward outcome measures, with outcomes taken as given, without recognition that the improvement of programs and persons comes only when outcomes (aims) are grounded in the desires of a specific group of people bound together in a common cause is likely to produce only the appearance of compliance without its substance. As Johnson (1989) asserts, 'The most effective evaluation for learning is self-evaluation' (p. 523), for only the results of self-evaluation reach down to what matters most to individuals in their professional lives and practice and stick.

Powell suggests there are three conflicting evaluation 'grammars', each embedded in a different set of assumptions about human nature, motivation, and learning and in different views of the location of legitimate authority. The first, 'self-review', is closely akin to the model developed by NCATE for teacher education accreditation. The second, 'self-evaluation', is quite close to the approach taken by TEAC (www/teac.org), a small but growing competitor to NCATE, which emphasizes program auditing. More will be said about TEAC later. The third she describes as 'self-assessment', which is close to the model under development at Brigham Young University, which will be explored in some detail shortly. While these terms are often used interchangeably, she argues they and their effects are radically different. She characterizes the differences among the three grammars in the following terms, which capture differences in the way in which power is distributed. (This point is important because of the relationship between power and opportunities to learn and to develop.)

Grammar I 'combines imposed external accountability, under the illusion of self-assessment, with a reduction in the cost of external inspection and audit. Self-review can be viewed as another form of control, as practitioners' ability to influence improvement and development is extremely limited' (Powell, 2000, p. 41). This is so because the outcomes, like those developed by NCATE, are established externally and what counts as

data in support of having met the outcomes is highly circumscribed. Moreover, the focus on the institution or unit, rather than program or person, diminishes the role of individual development in the story of the achievement of external aims.

Grammar II is 'more participative and teachers are regarded as an important source of information. Improvement is incremental and usually small scale. As with [Grammar I], the organization is the predominant definer and originator of change. The capacity of individual teachers to shape and sustain significant change processes is extremely limited' (p. 41).

Grammar III 'presents self-assessment as a process-led collaborative enterprise. The demand for greater holism emphasizes the strategic role exercised by the individual teacher. Through the dialectic, individuals are aligned to a sense of common purpose, committed to change in themselves and in their organizations' (p. 41). Powell goes on to suggest: 'Self-assessment must involve teachers from all levels in the school in the regular and systematic review of their processes and results' (p. 42). In short, everyone must participate in self-improvement efforts with the result that what emerges is a culture of assessment, a 'culture of continuous improvement', supported by relationships that sustain individual learning: 'If schools aim to improve continuously, then to achieve this, change can no longer be the exclusive responsibility of school managers, but must energize the talents of all teachers school-wide' (p. 42). The focus on the development of individuals within a culture designed to facilitate learning is central to the approach to evaluation unfolding at Brigham Young University.

Accreditation and a Clashing of Grammars

The story of the recent NCATE review of Brigham Young University during the spring of 2002 illustrates the weaknesses in the Grammar I approach to accreditation embraced by NCATE, and underscores the difficulty of challenging the deep technocratic bias inherent in the review process itself. Through our contacts with other teacher education institutions in the USA, we have come to believe that our experience is not atypical.

TABLE 1. The three evaluation grammars

Grammar I	Grammar II	Grammar III
Self-review	Self-evaluation	Self-assessment
Controlling	Improving	Creative
Retrospective	Incremental	Dynamic change
Mechanistic	Participative	Holistic
Summative	Limited feedback	Formative
Teachers as resource	Little individual ownership	Teacher role strategic
Not developmental	Improvement	Improvement
Outcomes	Mainly outcomes-led	Process-led
Routine	Power still centralized	Collegial multiple stake holders

BRIGHAM YOUNG UNIVERSITY: PREPARING FOR REVIEW

BYU is one of the largest teacher education certification institutions in the US, graduating approximately a thousand intending teachers per year. Among the distinctive features of BYU is its involvement as a founding member of the National Network for Educational Renewal and its support of perhaps the largest school/university partnership in the nation (see Goodlad, 1994). Last accredited by NCATE in 1997, BYU fell within the transition period, and organized accordingly. BYU has a large teacher education unit comprising several programs and some 40 separate departments, some of which have their own teacher preparation programs, with their own guiding conceptual frameworks (e.g. Music Education, Audiology and Speech Pathology as well as various graduate programs).

Recognizing the complexity of program evaluation and anticipating a shift to performance over input measures for accreditation, work to meet Standard II, on assessment, began in earnest in 1998. Some effort went into updating the conceptual framework from the 1997 review. NCATE standards clearly allow multiple frameworks and the university has a clear and direct mission statement that unites the various programs across campus in a shared institutional vision. Data routinely collected by the university and departments on student performance and program quality were reviewed. Plans were then made to organize these data for the visiting NCATE team. Data were available on student grade point averages, test score averages of various kinds, courses of study, graduation and employment rates as well as various demographic measures (gender, ethnicity, age, and so on). These data demonstrated conclusively that the teacher certification graduates of BYU are, like the student body generally, exceptionally able students academically, students whose performance on standardized tests is comparable to the best colleges and universities in the US. If, as some teacher education detractors argue, academic ability and scores on standardized achievement tests are good measures of teacher quality, these students ought to be outstanding.

Data were available from some three dozen program studies that had been conducted over the past few years each of which in one way or another related to quality issues. These studies were reviewed. An effort was made to identify gaps in the existing data sets, and plans were made to fill the gaps. In particular, there was concern about gathering data that would show what exactly was taught in the various courses and the value to the students of what was taught. The phrase ‘value added’ began to be tossed around in faculty meetings. Plans were made and implemented so that every elementary teacher education student created a portfolio that represented what they knew and were able to do. While promising, it became apparent that the amount of data contained in the portfolios was overwhelming and that to do an adequate analysis of hundreds of portfolios would require tremendous and unavailable resources. Nevertheless, the value of the portfolios to candidate growth was apparent although how precisely to use these data to help build the case for unit quality was a task set for the future.

Rethinking Assessment

As this work progressed, the professor (trained in educational psychology and program evaluation) assigned full-time to lead the effort reviewed the development and implementation of assessment programs at other US institutions that had received praise for their work. It is worth noting that these were, in contrast to BYU, primarily small state-supported colleges with central administrations much involved in the assessment movement and in direct control of their programs. We turn now to the professor's own words to describe what transpired:

> Impressed by these efforts, I developed an elegant structure and distributed it to the faculty for a proof-of-concept. Because the plan was tedious to implement, and because assessment was low on faculty priority lists, the plan was largely ignored. This lack of willingness prompted me to go into offices and engage in one-on-one conversations about assessment beliefs of the faculty, priorities, and the way they studied their own teaching. Conversations and subsequent faculty meetings led me [to reconsider my model] and led me towards a more realistic, responsive and functionally relevant assessment approach, one that encouraged the overall study of one's practice.

What followed was a fundamental reconsideration of assessment within the McKay School of Education. In effect, a shift began to take place from a Grammar I to a Grammar III approach to evaluation, from a top–down and externally imposed plan to one that emphasized self-assessment—conversation, communication, and decision-making driven by internally generated data—and culture building. This was done in the belief that the NCATE transition period would allow sufficient time to generate the sort of evidence needed to make a compelling case that the programs added value.

It was believed that this shift would produce genuine and sustainable program improvement. Genuine because faculty members would be working to improve their own practice which directly affected students, and sustainable because improvement would be internally motivated, driven by deep and personal faculty member interests rather than imposed from without. Following this shift in orientation toward assessment, numerous faculty members gradually became involved in the study of their practice. Simultaneously, syllabuses were gathered for every course offered and a study was made to determine underlying aims of each course and to determine what evidence was available that supported accomplishment of these aims. A series of meetings followed with several faculty members to assist them to better articulate aims and means and to assist in data gathering efforts. The fruit of this labor came in the form of the improvement of individual courses, which, it was believed, would improve in turn the quality of the programs. In Powell's terms, self-assessment was given priority over self-review and evaluation.

As the time for the NCATE visit neared, the institutional report was written which included a description of the alternative approach to assessment that was unfolding along with a timeline for further development. Data of various kinds were included that supported the claim that BYU graduates are quality beginning teachers. For example, entry

and exit data were provided on elementary program applicants, including evaluations of student teaching. Importantly, data from the five partnership school districts on beginning teacher performance were also included. The faculty realized that much work remained to be done, but most faculty members were generally confident that the direction taken was promising and that better programs would result because of the emphasis on faculty involvement in and responsibility for evaluation and self-assessment. Faculty leadership and several faculty members understood clearly that building an assessment culture would take time and change would be incremental but long-lasting.

The NCATE Board of Evaluation Arrives on Campus

As the visit unfolded, faculty concern began to rise. Two years prior to the visit and in preparation for the review, the university coordinator for the review along with the department chair and associate chair of the Teacher Education Department attended NCATE 2000, Continuing Accreditation and Beyond, as well as additional national meetings where NCATE officials explained the evolving standards and forthcoming changes in procedures. Another associate chair had 10 years of experience with NCATE as a professional association representative and clearly understood the processes and intent of the review. The message given at these meetings proved to be inconsistent with what eventually transpired in the review. Looking to the future, in the meetings heavy emphasis was placed on preparing Performance Assessment Plans (PAPS). Meetings were fraught with complaints; a general feeling of uncertainty persisted. Representatives of sponsoring professional associations that were to participate in specific area reviews noted that NCATE was sending them a different or no message at all. Finally, the BYU group was given a document from NCATE dated 10 July 2000 from which to prepare for the upcoming review. Unfortunately, the visiting team used a different document, one dated a year later. The revised document included details related to the standards not present in the first document. Moreover, there were clear differences in expectations. But these problems were not of most concern. More serious issues were at stake that point to a fundamental weakness in the very foundation of accreditation.

In the light of their careful study of the NCATE handbooks and participation in various NCATE-sponsored meetings the BYU faculty members who prepared for the visit made some assumptions about the review that proved to be false. (1) They believed the NCATE handbooks when they stated that 'There are many different ways to meet NCATE standards. The way that they are met is often dependent on the mission of the institution and unit. Thus, [review] team members must guard against applying independent measures of quality that may be based on their experiences in various research settings: religious, historically black, private, state, or land grant. Familiarity with one's own approach to addressing standards sometimes limits the ability to accept other approaches as valid' (NCATE, n.d., p. 47). (2) As already noted, based upon NCATE documents, they thought it was possible for a single institution to have multiple and differing conceptual frameworks but still pass accreditation. And (3) they understood the transition period was in effect and full implementation of assessment plans was not expected: 'Units with visits scheduled fall 2002 and spring 2002 should have a well-developed plan for an assessment

system that includes timelines and details about the system components. The plan should have been developed collaboratively by members of the professional community …' (NCATE, 2000, p. 16). As the hours passed, it became increasingly clear that having worked from these assumptions to prepare for the visit was a mistake and that passing Standard II was potentially in jeopardy.

A Worried Response

A meeting took place between the faculty member assigned to head up the response to Standard II (assessment), the area that seemed most vulnerable, and two members of the review team. The intention was to better explain the approach being taken to assessment and to answer any questions team members might have about the approach. We turn to a document written by the professor in charge of the response to Standard II for insight into what transpired in this meeting:

> I handed the two team members my written document titled, 'The Study of Practice and Programs' (which, unfortunately did not have a sub-title, 'A Unit Assessment Plan'). This document represented progress to date and contained specific recommendations for engaging in a study of one's teaching and the group's study of our program. The document was set aside. Both visitors were visibly restless, [uncomfortable] with the direction I was taking them and they did not want to hear much more. Taking control of the conversation, they wanted to see a diagrammatic picture, a flow chart showing the elements and timelines of a 'standard' unit assessment plan. My approach was incompatible with their perceptions of how a plan should be developed. I had taken them on a journey rather than presented them with [something they could recognize as] an assessment plan. The person who had started out reassuring me [that NCATE was in transition] now shifted to counselling me about re-adjusting my thinking so as to end up fitting NCATE guidelines. Giving them the benefit of the doubt, I backtracked and re-explained that my initial work had been to do just that but the result was unacceptable to colleagues and therefore non-implementable.
>
> The real agenda began to emerge. The second team member entered the discussion with a statement something like, 'The faculty we have interviewed keep telling us they've been gathering data all along but I can only see bits and pieces here and there. Your data have not been gathered into one place and summarized to yield an overall picture of the effectiveness of your program. Do you have any examples of how you have used data to make programmatic changes? I want to see data'. Clearly, each visitor had a different purpose. One wanted specific commitments about (1) phases of development and implementation over the next 5 years, and (2) an assurance that the finished model would fit that prescribed NCATE guideline. The other, more impatient, wanted to see data at this point in our development. Both left feeling they had accomplished the purpose of their visit to my office. Because these team members were template-oriented, we predictably failed Standard II. Both were from small state-supported colleges where the assessment office is centralized and under the direct control of the administration. Above all was their

> attempt to expound the virtues of the systems in their own institutions. They failed to realize (or their role prevented it) they were dealing with a large university where top–down control of assessment would never work either politically or functionally.
>
> The implicit control showing through would be resisted at every turn. The contexts were too different. Still, they had been given a set of marching orders by NCATE. (Contrary to Grammar III values) their judgments would be based on the degree to which our assessment activities fitted their template.

The Examiners Report

The Board of Examiners concluded that BYU passed five of the six standards, but not Standard II. The description contained in the Board of Examiners Report of why the unit failed Standard II is revealing. A few representative quotes from the report follow along with a our commentary:

> The narrative in the Institutional Report, the subsequent documentation available, and interview responses did not explain how current, available data collected is systematically aggregated …
>
> Upon completion of [the] program, final student teaching assessments in narrative form are collected, but this data set varies between programs within the unit, and the information is not currently aggregated to provide unit-wide data.

This is a Grammar I assertion, tied to the concern with 'units', that represents a view of assessment that ignores program variation (of which there is necessarily a great deal) in both aims and means. Apparently, while variation in conceptual frameworks is in principle possible, in practice, when it comes to assessment, it is to be discouraged (so wise faculty will hide the reality of diversity when being reviewed and in written self-reports). Moreover, the demand for data in a period of transition goes well beyond the remit of the team. Continuing: the focus of NCATE on 'units' results in a serious and hidden bias against large and complex teacher education institutions where governance is widely shared and responsibilities distributed broadly. From a faculty point of view, 'unit' is a meaningless abstraction. Faculty members work in and seek to improve courses and programs, not units. More importantly, BYU has developed a Center of Pedagogy (see Patterson *et al*., 1999) that places governance of teacher education in the hands of representatives from the arts and sciences, the school of education, and the five partnership school districts. Beneath this umbrella, which coordinates a range of complex, diverse, and far-ranging teacher education efforts, stand numerous, relatively independent, programs each of which is in some sense a 'unit'. Partnership among the three central interests in teacher education—arts and sciences, schools, and colleges and departments of education—which is a precondition for quality teacher education, brings with it multiplicity of aims and means. Unfortunately, NCATE standards have difficulty encompassing something so complex, an organization that defies simple, direct, and unidirectional lines of authority, and is ever-changing.

'Multiple interviews with primary key stakeholders for assessment yielded the reality that a unit-wide assessment system plan, one designed, planned, and built on consensus among the professional community that reflects its conceptual framework and professional and state standards for the unit as a whole, was not available.' Given the size of the 'unit', the range of programs represented within it, and the diverse interests of members of the BYU partnership, there must be multiple evaluation systems in place. Most importantly, however, so deep were the differences in understanding of the nature of evaluation, when meeting with the faculty member in charge of the assessment component of the review and being handed the plan, it was not recognized as even being a plan! Grammar I eyes were unable to recognize Grammar III visions. 'As a result of the diverse data sources, the unit makes limited or no use of data collected …' As noted, the 'unit' is not the appropriate 'unit' for data gathering and analysis nor the best place to direct improvement efforts. As the BYU plan suggested, we believed the most productive focus for improving candidate competence was on programs and on individual instructors and their teaching practices. This is a Grammar III concern. Clearly, aggregation of data at the unit level has only very limited value to faculty growth and program improvement. Grammar I assumptions failed to produce results.

Several recommendations for improvement are offered in the Board of Examiners Report. Each recommendation reflects Grammar I assumptions about assessment, and is insensitive to the challenges of evaluation in a large and complex university on one hand and to an institution committed to school/university partnership on the other, where governance is genuinely and widely shared. One follows—tied closely to documenting performance outcomes—along with the 'rationale' offered for the recommendation: 'The unit assessment plan lacks an integrated set of measures used to provide information in mentoring candidate performance and managing and improving programs and unit operations'. The rationale: 'The unit has not identified what specific assessments will be used and when. Programs often act autonomously in regards to assessment, particularly the collection, analysis, and dissemination of assessment data. Faculty members are unaware of and unable to articulate the importance of a unit-wide assessment system designed to directly assess candidates' progress on learning outcomes identified in the conceptual framework (this is because they don't see much value). The plan articulated in the Institutional Report is limited in addressing faculty evaluation and unit operations.'

Clearly, each of the three operating assumptions proved false: there is a very narrow range of possible ways to meet the standards; there must be a single 'unit' vision and system; and there was no transition period. The importance of the later point is that the standards were inflexible. Standing back and considering the review, we find reason for grave concern: if the experience of BYU faculty during the review is repeated elsewhere, and we suspect it will be, will the effort to develop innovative approaches to assessment along Grammar III lines be undermined? We think so. Consider Powell's listing of assessment grammar attributes and the close fit of the concerns of the Board of Examiners with Grammar I: control ('validity', 'substantive detail', 'mid-point progression assessments', 'mid-point skills', 'consistent across all programs'), mechanism ('comprehensive measures'), summation ('unit', 'summarized', 'aggregation'), routine ('system', 'consensus').

Then contrast these with Grammar III attributes and their responsiveness to the dynamic and uncertain nature of teacher education.

Attempting to make sense of what happened in the review, the faculty member responsible for responding to Standard II nicely captured and extended the intent of Grammar III assessment and underscored the danger of assessments driven by Grammar I commitments:

> Carrying in mental templates is the unwitting extreme end of standardization. Authentic assessment, deep-running assessment, productive assessment, needs to be generated from within specific contexts—within the value structure and political milieu of each [teacher education] institution, a reality antithetical to standardization. Furthermore, these marching orders, well meaning as they might be, breed an unapproachable naiveté in the minds and dispositions of the messengers. When one carries in a standard template, listening and learning are irrelevant to the conversation. These team members did not have the freedom (or inclination) to maturely respond [by saying], 'You have an unusual approach to assessment of students and programs, a bit foreign to us. Still, can we bring to the table our experience that will help you develop it?'. To make such a statement violates the responsibility conferred upon them. To be fair, on the other hand, it must be said that the spirit behind these standards is reasonable and persuasive: encourage institutions to upgrade the quality of their assessments so that better teachers can be cultivated. When this gets down to the operational level, however, sending out teams with templates of certification literally insures the means becoming the ends.

What Now?

A rejoinder was prepared and submitted. Just as in the schools, the power of Grammar I assessment, and the technocratic mindedness underlying it, is overwhelming in teacher education accreditation. Standardized and simple measures are, from the perspective of Grammar I assessment, most valued. Complexity in the education of teachers is denied in an ill-fated quest for certainty and uniformity of outcomes. In this context, the burden of proof clearly rests on those teacher educators who would argue for program quality on Grammar III grounds. It may or may not have been the case that the examiners who visited BYU could have been convinced of the worth of the alternative approach to assessment being developed. What is certain is that the case that was presented was judged not fully adequate. To be sure, errors were made in the way in which the self-study report was organized and written. Some of these errors came from assuming the transition period allowed for reporting only the data currently available. More data could have been presented, and presented in more effective ways. The easy solution, as one long-time NCATE examiner recently commented as he described how he was approaching his own institution's immanent review, is to 'write the report just as NCATE wants', to fit the standard template. But fitting the standard may not be proof of quality nor will it aid in program improvement or promote faculty learning.

It is in part because of concerns of this kind that TEAC was formed. Although just beginning, the TEAC challenge to NCATE is promising. TEAC differs from NCATE's approach to accreditation in at least four ways. (1) TEAC provides coaches to assist faculty to prepare an 'Inquiry Brief' used to make the case for accreditation. Consider: BYU sent representatives to various NCATE training meetings but despite consistent effort to gain clarity on NCATE intentions, confusion and uncertainty persisted. (2) TEAC sharply focuses on programs and on candidate performance. Four of six NCATE standards require documentation well beyond candidate performance. The size of the 'unit' in large institutions like BYU and NCATE insistence that every aspect of a 'unit' be assessed (including graduate and undergraduate programs) makes it so that individual program quality gets lost in the shuffle. (3) TEAC conducts 'audits' which begin 'with a period in which the auditors essentially "play back" to the faculty their own understanding of the program's "story" [which] demonstrates auditors have read and understood the brief' (Ewell, 2001, p. 15). During the BYU visit, there was a strong sense that the accrediting team did not have a clear or full picture of the institution or of its aims. In particular, there seemed to be little understanding of the role of the Center of Pedagogy (CITES—Center for the Improvement of Teacher Education and Schooling) in governance and in bringing together the disparate bodies involved in teacher education to strengthen programs, overcome differences of opinion, and to engage in what John Goodlad calls 'simultaneous renewal'. (4) As the review process unfolds, TEAC gives 'Programs … the chance to supply additional information, change the nature of what was being claimed or to withdraw certain claims or statements entirely' and to include a wider range of evidence than is typically allowed (Ewell, 2001, p. 15). NCATE reports are prepared months in advance and are set by the time the Board of Examiners arrives. The result is that misunderstandings are extremely difficult to correct. This is not surprising given the scope of the team charge, to assess virtually everything about a 'unit'.

At this point, making a non-standard case for program quality is the challenge facing BYU and many other teacher education institutions whose aims and programs challenge established prescriptions. If a successful case is to be made, it will require a tremendous increase in resources dedicated to program evaluation. Already, NCATE accreditation is unbelievably expensive. It will involve serious thought about what counts as good evidence and how to gather it. And it will involve significant political and cultural work, because ultimately good programs are a function of the characteristics and qualities of diverse and complex human relationships and of institutional and personal commitments. Part of the political work needing to be done will be with NCATE itself. Ultimately, the quality of a teacher education program is a reflection of the state of the hearts and minds of teacher educators and of their desire and ability to imagine their work in new and refreshing ways and then to take concerted action to realize their visions. Fear of diversity, distrust of teachers and teacher educators, a naive faith in the ability and value of systems to control behavior and to assure quality performance, and narrow conceptions of the nature of teaching and learning to teach conspire to produce template-bounded minds. Learning to live, in the words of one struggling teacher, within a 'straightjacket of outcomes' (Arey, 2002, p. 33) may produce minds that reside comfortably within bounds,

but not good teachers. Nor can Grammar I accreditation systems produce quality teacher education.

REFERENCES

AREY, K. (2002) When standardization replaces innovation, *Education Week*, 21(34), pp. 33–34.

BULLOUGH, R.V., JR., GOLSTEIN, S.L. & HOLT, L. (1984) *Human Interests in the Curriculum: Teaching and Learning in a Technological Society* (New York, Teachers College Press).

COCHRAN-SMITH, M. (2001) The outcomes question in teacher education, *Teaching & Teacher Education*, 17(5), pp. 527–546.

EWELL, P.T. (2001) *Piloting a New Approach to Accreditation in Teacher Education: An Evaluation of the TEAC/FIPSE Project* (Boulder, CO, National Center for Higher Education Management Systems).

GALLAGHER, J.J. (2002) What next for OERI?, *Education Week*, 21(28), pp. 52, 38.

GILROY, P. (1999) Inspecting the inspecting of teacher education in England and Wales, *Journal of Education for Teaching*, 25(3), pp. 215–219.

GOODLAD, J.I. (1994) *Educational Renewal: Better Teachers, Better Schools* (San Francisco, CA, Jossey-Bass).

HOFF, D.J. (2002) Md. to phase out innovative testing program, *Education Week*, 21(29), pp. 27, 29.

JOHNSTON, P. (1989) Constructive evaluation and the improvement of teaching and learning, *Teachers College Record*, 90(4), pp. 509–528.

KELLER, B. (2002) Michigan on verge of getting new system to grade schools, *Education Week*, 21(28), p. 21.

LINN, R.L., BAKER, E.L. & BETEBENNER, D.W. (2002) Accountability systems: implications of requirements of the No Child Left Behind Act of 2001, *Educational Research*, 31(5), pp. 3–16.

NATIONAL COMMISSION ON EXCELLENCE IN EDUCATION (1983) *A Nation at Risk* (Washington DC, US Government Printing Office).

NCATE (n.d.) *Conducting the Continuing Accreditation Visit: Guidelines for Board of Examiners Teams* (Washington DC, NCATE).

NCATE (2000) *Handbook for Continuing Accreditation Visits* (Washington DC, NCATE).

NCATE (2002) *Professional Standards for the Accreditation of Schools, Colleges, and Departments of Education* (Washington DC, NCATE).

NELSON, T. (2002) Editor's introduction: teaching for success, *Teacher Education Quarterly*, 29(3), pp. 3–5.

PATTERSON, R.S., MICHELLI, N.M. & PACHECO, A. (1999) *Centers of Pedagogy: New Structures for Educational Renewal* (San Francisco, CA, Jossey-Bass).

POPPLETON, P. (1999) Leading from behind: a comparative view of the proposed reform of teacher training, *Journal of Education for Teaching*, 25(3), pp. 233–245.

POWELL, L.A. (2000) Realising the value of self-assessment: the influence of the business excellence model on teacher professionalism, *European Journal of Teacher Education*, 23(1), pp. 37–48.

SMITH, M.L. & FEY, P. (2000) Validity and accountability in high-stakes testing, *Journal of Teacher Education*, 51(5), pp. 334–344.

TYACK, D. & CUBAN, L. (1995) *Tinkering toward Utopia: A Century of Public School Reform* (Cambridge, MA, Harvard University Press).

UNIVERSITIES COUNCIL FOR THE EDUCATION OF TEACHERS (1999) Evidence to the British House of Common Select Committee on the Inspection of Teacher Education in England, Wales and Northern Ireland, *Journal of Education for Teaching*, 25(3), pp. 221–231.